

Multi-Port Switch for PCI Express® Designs

Jason Monroe

Engineering Director

Synopsys

Copyright © 2018 PCI-SIG® - All Rights Reserved

Disclaimer

<u>Presentation Disclaimer</u>: All opinions, judgments, recommendations, etc. that are presented herein are the opinions of the presenter of the material and do not necessarily reflect the opinions of the PCI-SIG®.

Copyright © 2018 PCI-SIG® - All Rights Reserved

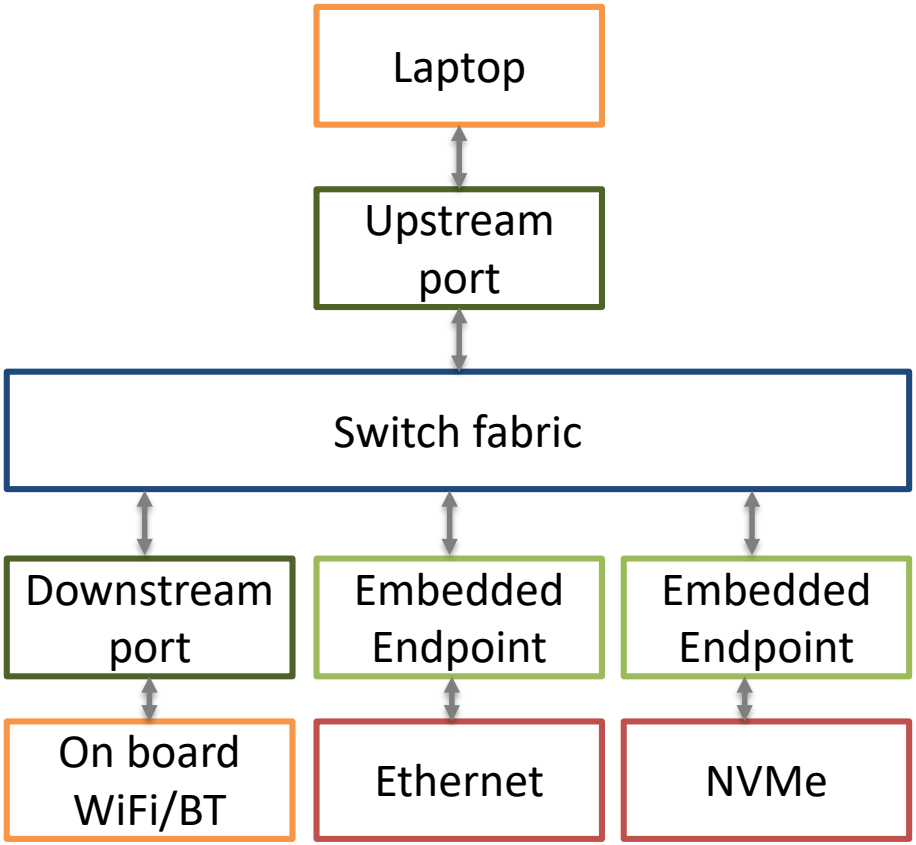
Agenda

- **Overview**
- **Architectural Highlights**
- **Example Data Flow**
- **Performance Metrics**

Copyright © 2018 PCI-SIG® - All Rights Reserved

Customer Requirements

- **Four port PCIe® switch for docking station**
- **Ports can be different number of lanes & speeds**



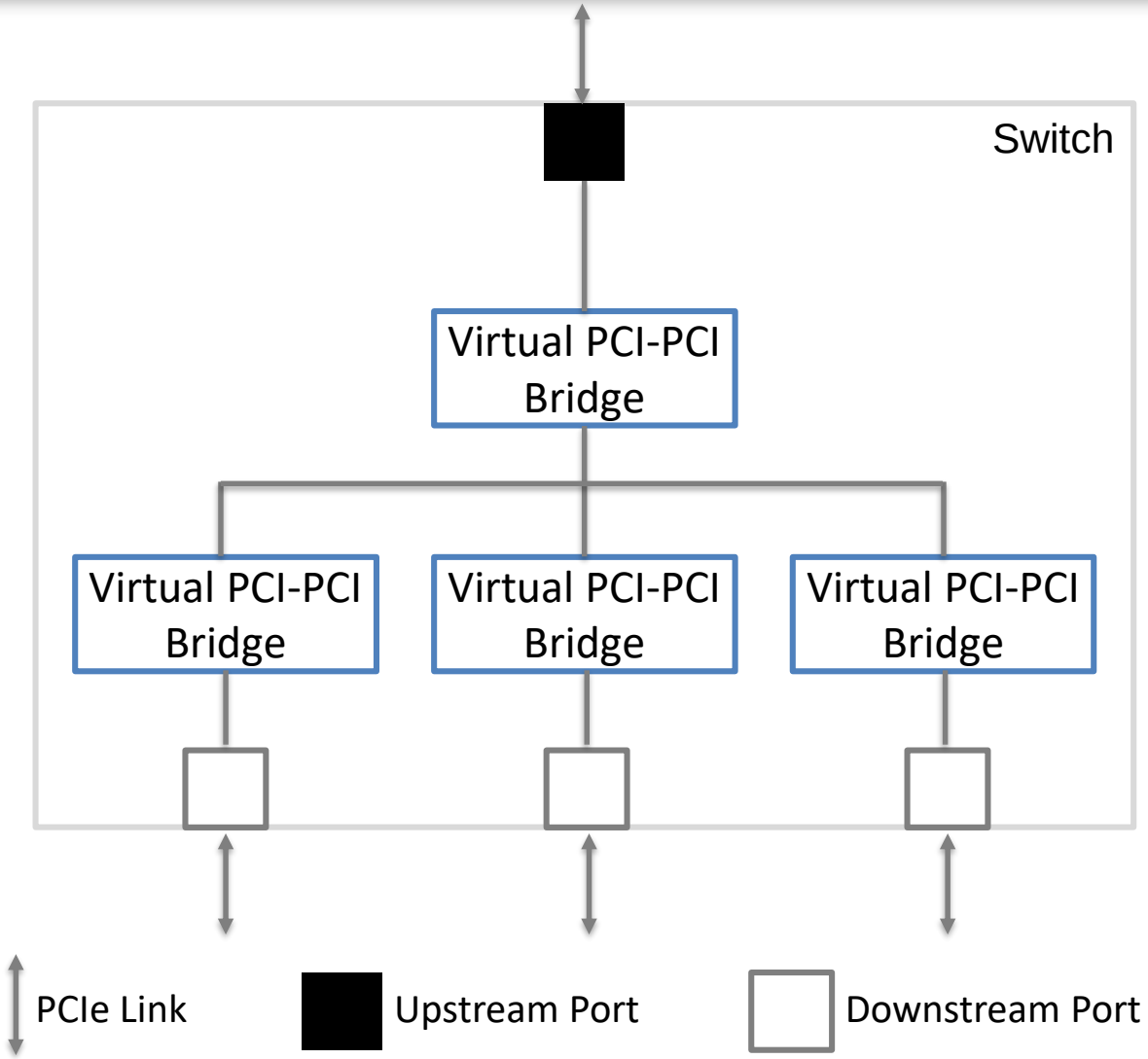
Copyright © 2018 PCI-SIG® - All Rights Reserved

PCIe Switch Design to Meet Your Needs

- **Scalable**
 - Multiple ports
 - Configurable Fabric bus widths and clocking
- **Reusable**
 - Create Methods to re-use Configurable Options Easily
 - Easily Integrates into SOCs
- **Low Overhead**
 - Maximize Throughput and Minimize Latency
- **Robust**
 - Interoperability is Important
 - Reuse Existing and Proven IP where possible

Copyright © 2018 PCI-SIG® - All Rights Reserved

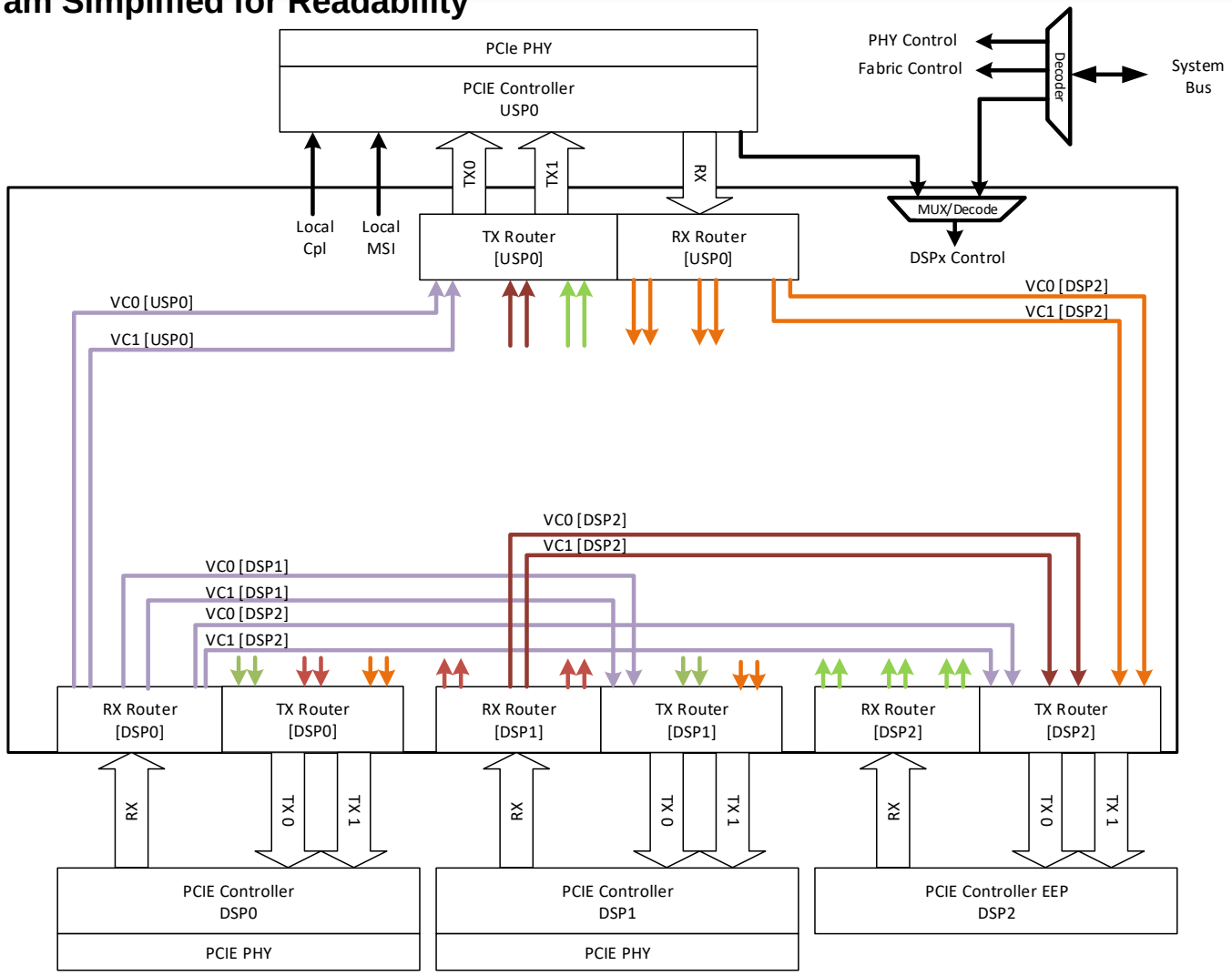
Logical Block Diagram



Copyright © 2018 PCI-SIG® - All Rights Reserved

High Level Architectural Diagram

- **Diagram Simplified for Readability**



Copyright © 2018 PCI-SIG® - All Rights Reserved

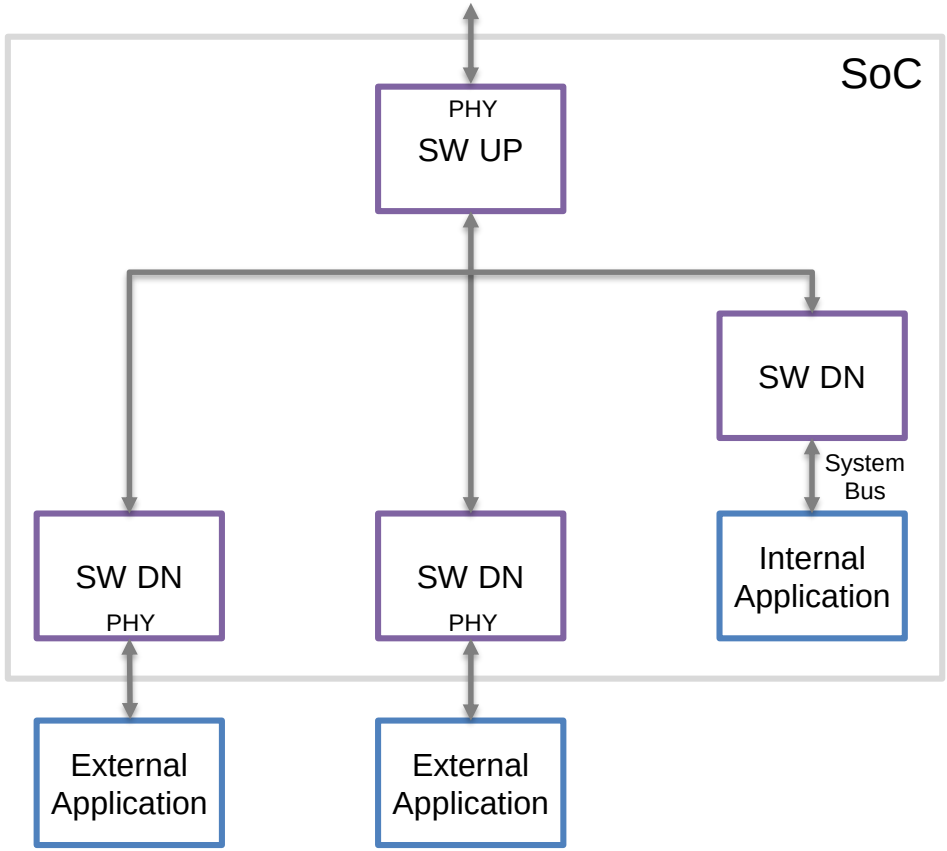
Switch Features

- **PCIe 1.0 → 5.0**
- **Up to 32 multiple downstream ports**
- **Peer to peer transactions**
- **Configurable bus fabric data width**
 - Trade off area vs. throughput
- **Support for virtual channels**
- **Cut-through buffering**
- **Designed to support all the required features of the PCIe specification**

Copyright © 2018 PCI-SIG® - All Rights Reserved

Need for Embedded EP

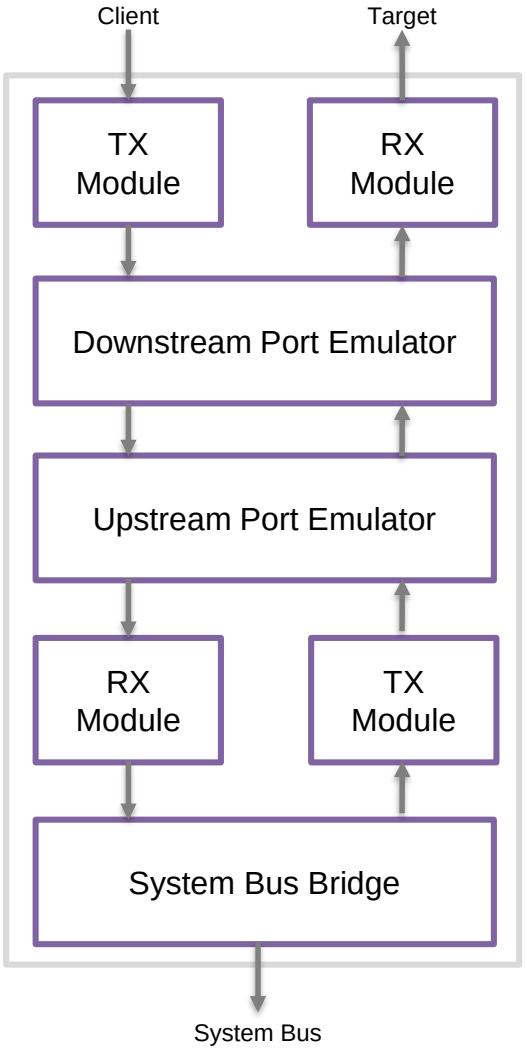
- **Switch fabric designed for SoC integration**
- **SoC may have internal EP function**
- **How to implement with traditional IP?**



Copyright © 2018 PCI-SIG® - All Rights Reserved

Embedded EP

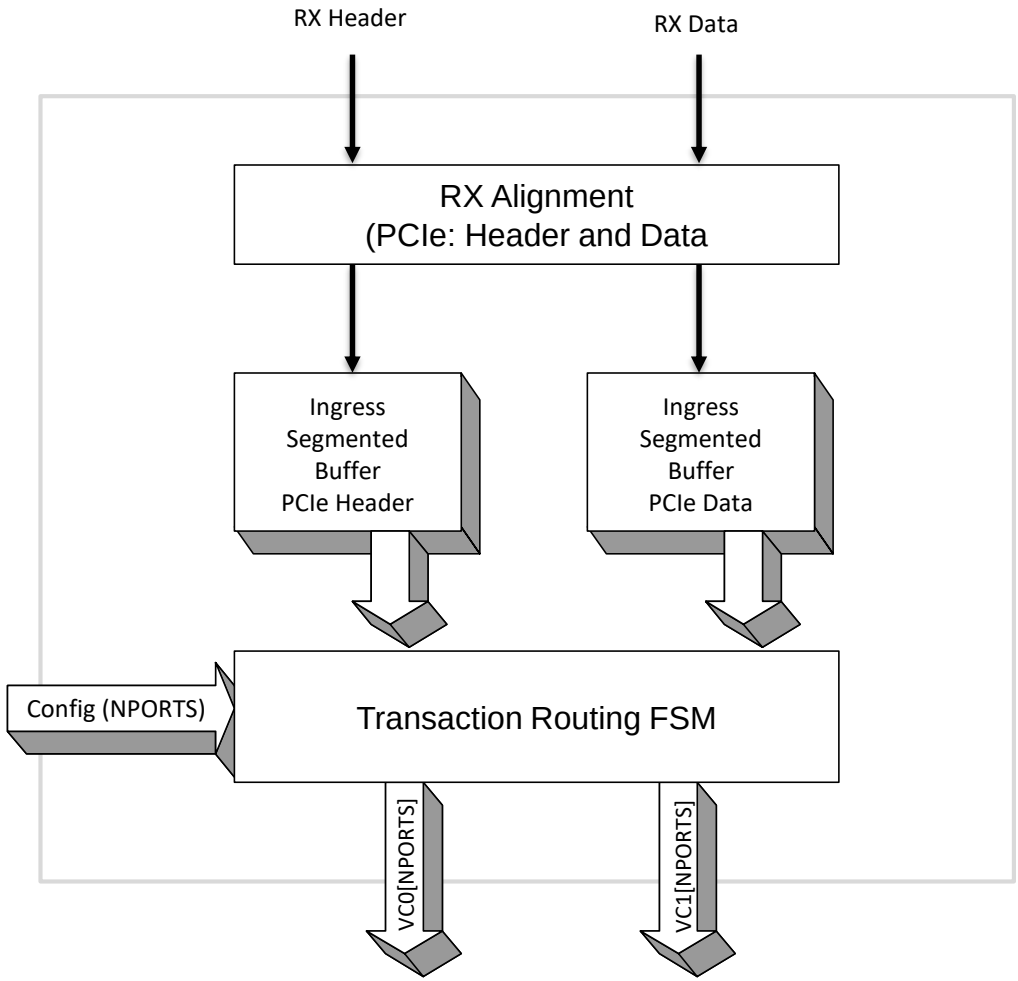
- **Merge downstream port and EP:**
 - Upper section is DSP switch port
 - Lower section is EP port
- **Fully compatible with PCIe hardware and software model**
- **No need for PHY or LTSSM**
- **No need to duplicate an additional link port**
- **Reduces total latency to the embedded application**
- **Reduces area and simplifies SoC connections**



Copyright © 2018 PCI-SIG® - All Rights Reserved

RX Routing

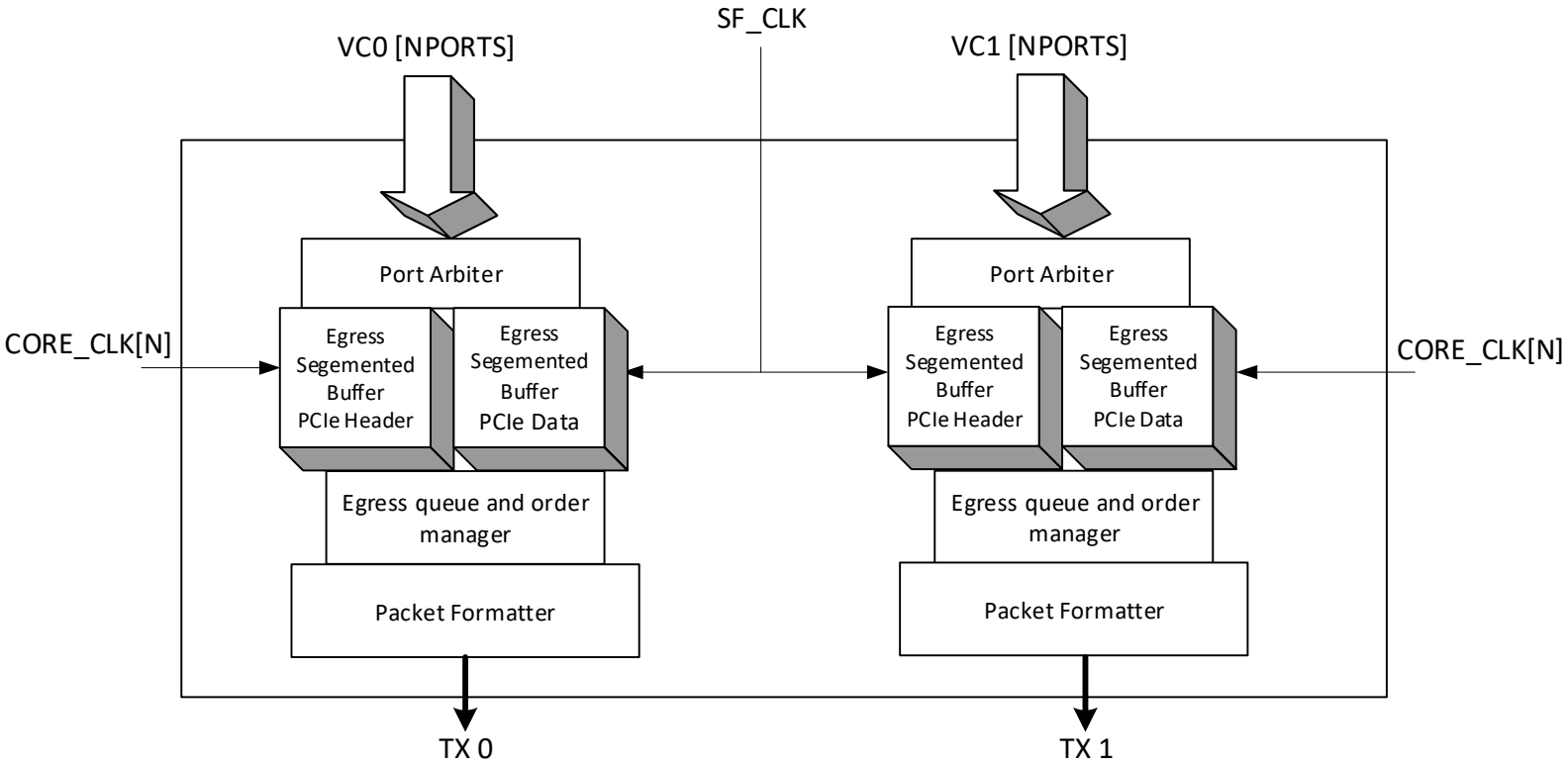
- **Programmable and Separate Buffers**
 - Supports Relaxed Ordering
 - Posted, Read, NPR with Data, Completion
 - Cut Through Buffers: Retain Data only long enough to determine route and not starve target port
 - Map Traffic to VC0 or VC1



Copyright © 2018 PCI-SIG® - All Rights Reserved

TX Arbitration

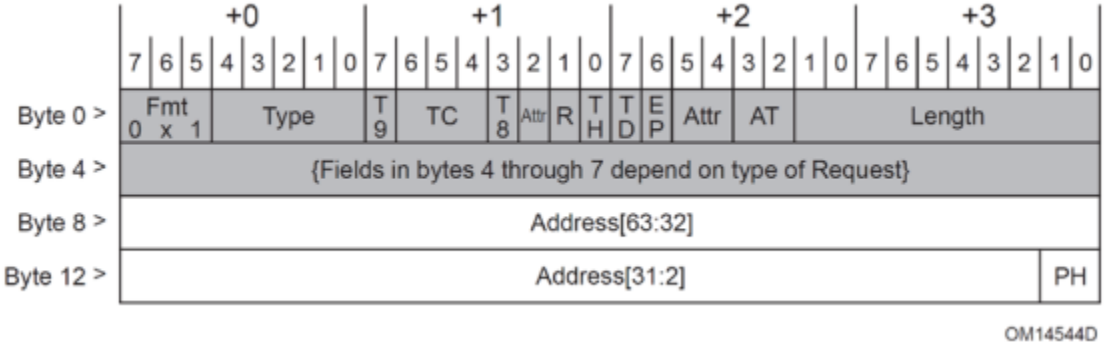
- **Maintain VC0 and VC1 Paths**
- **Arbitrate and Reconstruct Packets**

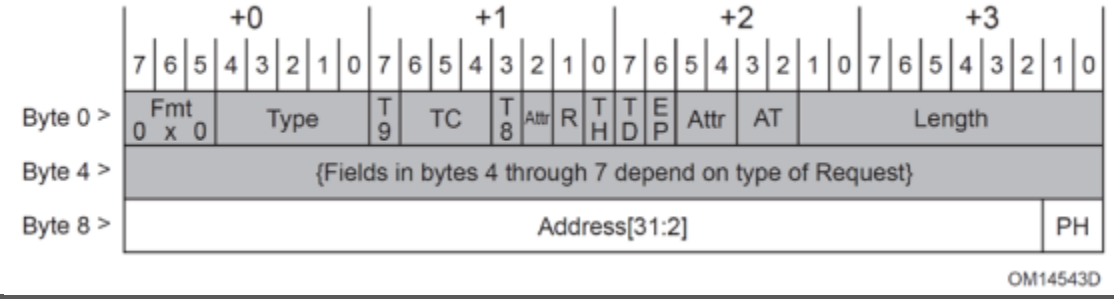


Copyright © 2018 PCI-SIG® - All Rights Reserved

How to Determine Route

- **TLP routing based on spec: Address, ID, and implicit**
 - Use Fmt and Type Fields
- **Address based (Memory and I/O):**

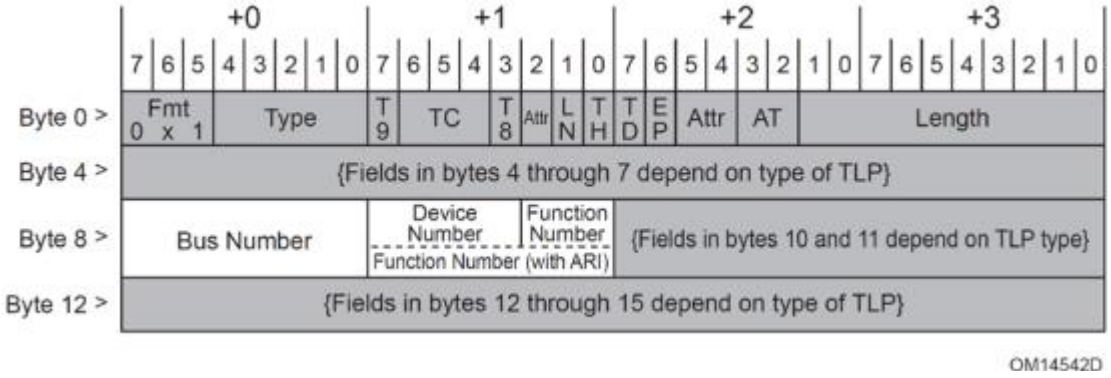


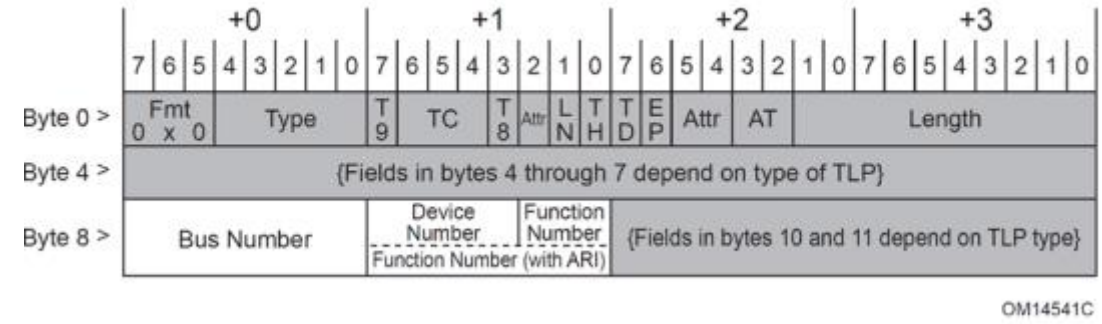


Copyright © 2018 PCI-SIG® - All Rights Reserved

How To Determine Route

- **ID Based**
 - Configuration Requests, ID Routed Messages, Completions





Copyright © 2018 PCI-SIG® - All Rights Reserved

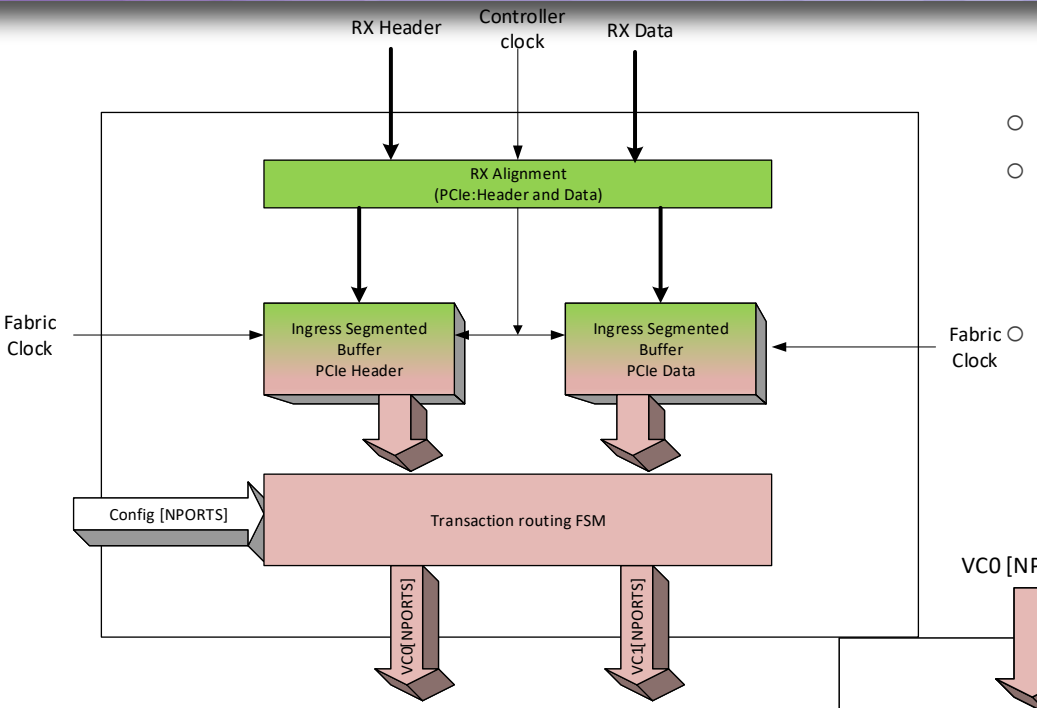
How to Determine Route

- **Implicit Routing**
 - Destination implied by routing type

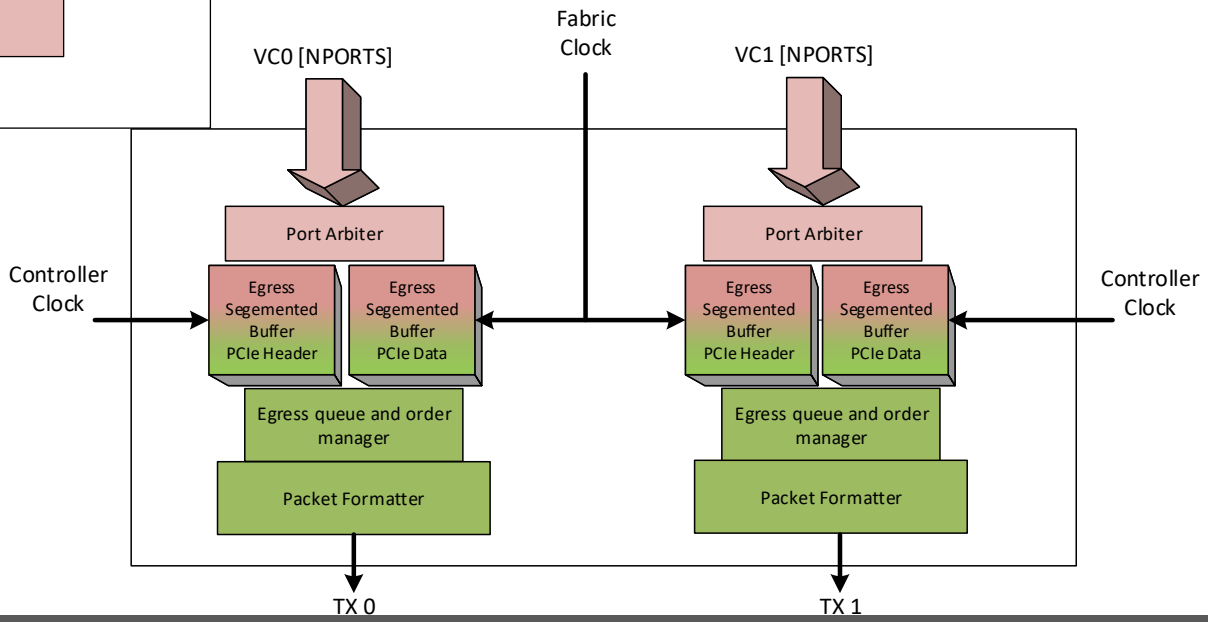
r[2:0] (b)	Description	Bytes 8 to 15[14]
000	Routed to Root Complex	Reserved
001	Routed by Address[15]	Address
010	Routed by ID	See Section 2.2.4.2
011	Broadcast from Root Complex	Reserved
100	Local - Terminate at Receiver	Reserved
101	Gathered and routed to Root Complex[16]	Reserved
110 to111	Reserved - Terminate at Receiver	Reserved

Copyright © 2018 PCI-SIG® - All Rights Reserved

Clock Domain Handoff

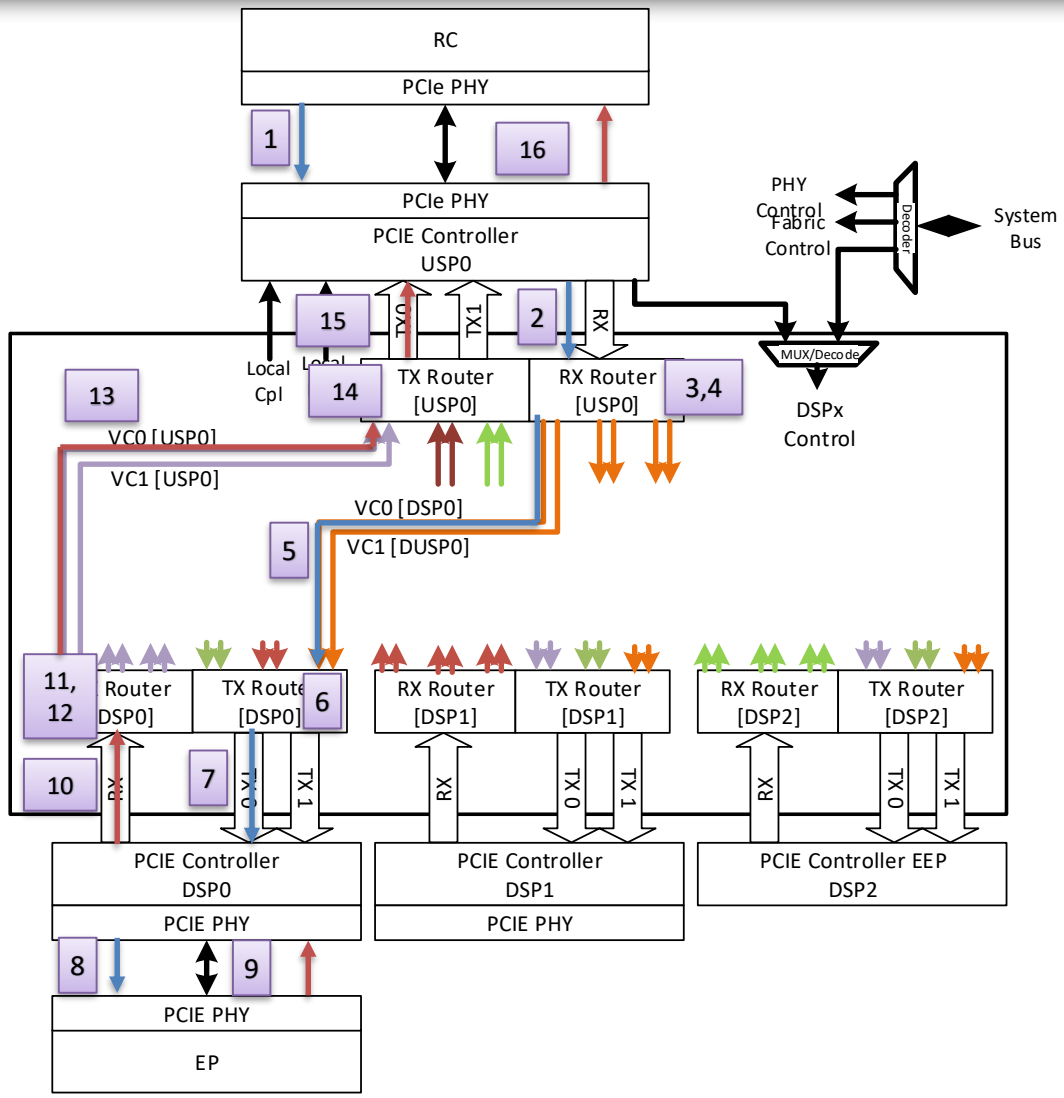


- **Fabric handoff through packet buffers**
- **Options**
 - Fabric clock is fully async
 - Fabric clock is synchronous to USP
 - Fabric clock is synchronous to USP and DSPs
- **Fabric clock selection dependent on total throughput required**



Copyright © 2018 PCI-SIG® - All Rights Reserved

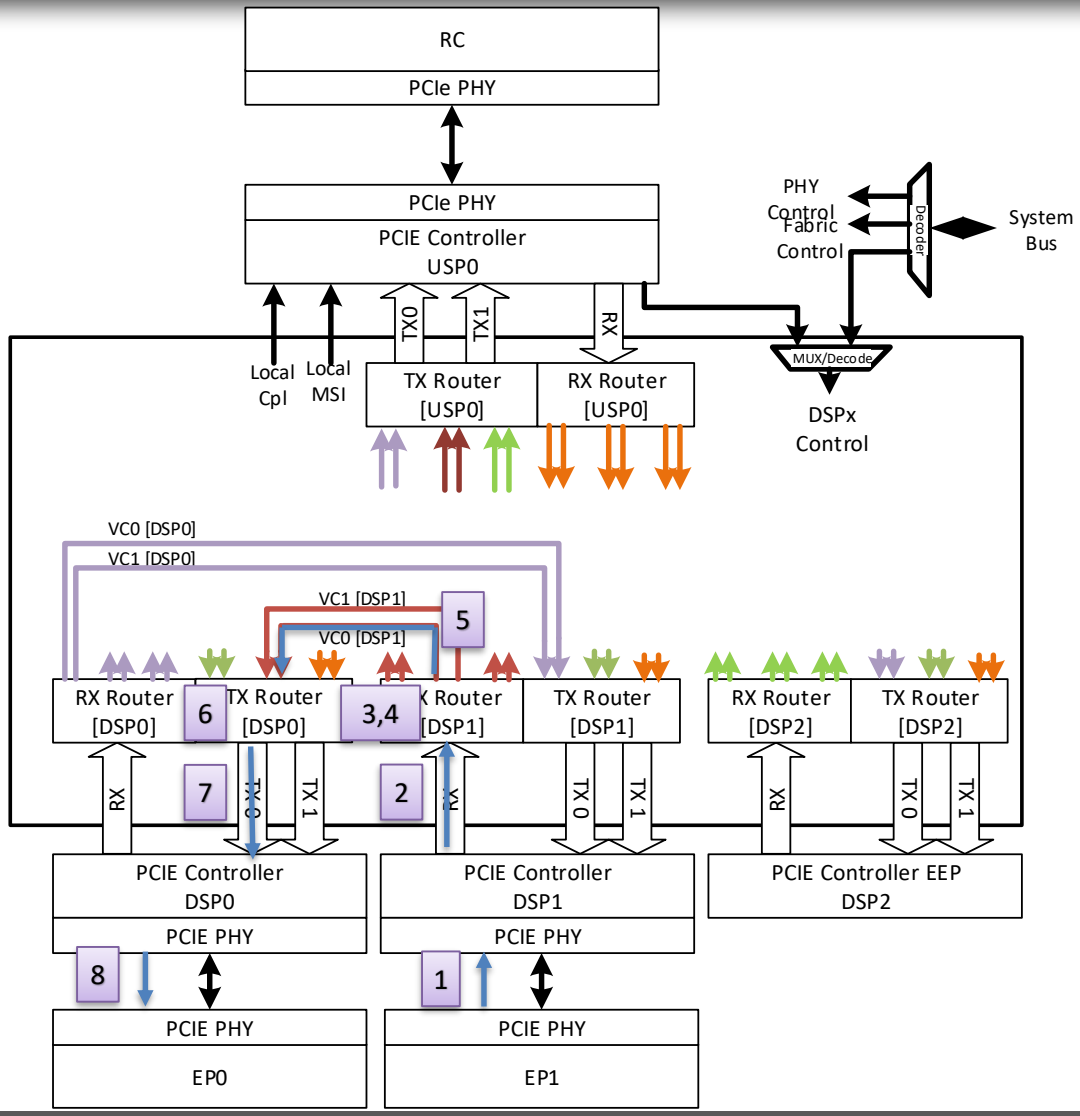
Example Data Flow MRd



1. RC sends MRd
2. USP0 Decodes request and forwards to switch
3. Rx Router USP0 stores request in non-posted queue
4. Following ordering rules, Rx Router USP0 reviews TLP header and config info from each DSP to determine route
5. Based on route and VC mapping, Rx Router USP0 forwards Mrd TLP to Tx Router DSP0.
6. After arbitration, Tx Router DSP0 stores MRd TLP in egress non-posted queue for VC0
7. Tx Router DSP0 pushes MRd TLP to DSP0 on TX0 (VC0) if space available.
8. DSP0 send MRd to EP
9. Assuming the transfer is valid, EP send CplD
10. DSP0 decodes CplD and forwards to switch
11. Rx Router DSP0 stores CplD TLP in ingress completion queue
12. Following Transaction Ordering Rules, Rx Router DSP0 reviews CplD TLP header and config info from USP and DSPs to determine route
13. Based on route and VC mapping, Rx Router DSP0 forwards CplD TLP to Tx Router USP0
14. After arbitration, Tx Router USP0 stores CplD TLP in egress completion queue for VC0
15. Tx Router USP0 pushes MRd TLP to USP0 on TX0 (VC0) if space available.
16. USP0 send CplD to RC

Copyright © 2018 PCI-SIG® - All Rights Reserved

Example Data Flow P2P MWr



1. **EP1 sends MWr**
2. **DSP1 Decodes request and forwards to switch**
3. **Rx Router DSP1 stores MWr TLP in ingress posted queue**
4. **Following Transaction Ordering Rules, Rx Router DSP1 reviews MWr TLP header (address based route) and config info from USP and DSPs to determine route**
5. **Based on route and VC mapping, Rx Router DSP1 forwards MWr TLP to Tx Router DSP0.**
6. **After arbitration, Tx Router DSP0 stores MWr TLP in egress posted queue for VC0**
7. **Tx Router DSP0 pushes MWr TLP to DSP0 on TX0 (VC0) if space available.**
8. **DSP0 send MWr to EP**

Copyright © 2018 PCI-SIG® - All Rights Reserved

Performance Metrics: Throughput

- **Typical throughput numbers**
- **All ports configured identically in a simulation**
- **Loss of throughput caused by protocol overhead and switch latency**

Profile	Gbps
PCIe 4.0 x2, 64-bit addr, MTU256, Write	28.3
PCIe 4.0 x2, 64-bit addr, MTU256, Read	27.2
PCIe 3.0 x1, 64 bit addr, MTU256, Write	7.1
PCIe 3.0 x1, 64 bit addr, MTU256, Read	6.8

Copyright © 2018 PCI-SIG® - All Rights Reserved

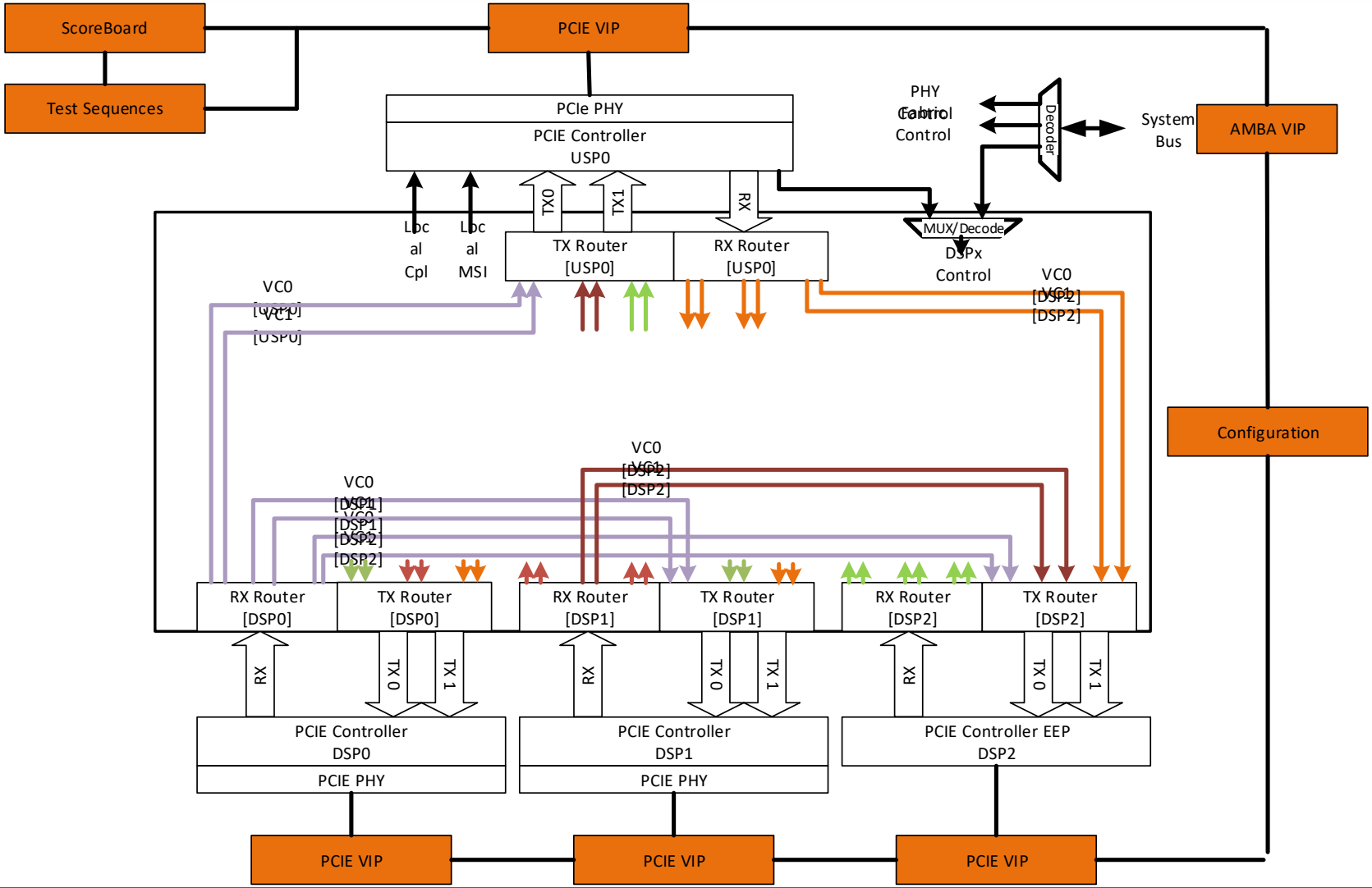
Performance Metrics: Latency

- **Latency added in the switch due to clock handoffs and routing**
- **Latency changes depending on traffic congestion and different port speeds**

Description	Cycles
Fully Synchronous mode	8
Fabric Synchronous to only DSP0	~10
Fabric Asynchronous to all port clocks	~12

Copyright © 2018 PCI-SIG® - All Rights Reserved

UVM Environment, Constraint Configurations and Traffic



Copyright © 2018 PCI-SIG® - All Rights Reserved

Summary

- **Users need a PCIe switch that is**
 - Scalable through configuration for clocking, number of ports, etc.
 - Reusable to address multiple technologies and applications
 - Optimized for low latency and area
 - Based on proven IP
- **Technologies like Embedded Endpoint Further reduces area and latency for applications embedded in an SoC**

Copyright © 2018 PCI-SIG® - All Rights Reserved

Thank you for attending the PCI-SIG Developers Conference 2018.

For more information, please go to www.pcisig.com

Don't forget to submit your feedback via the mobile app!

Download the **Crowd Compass** app and then search for **PCI-SIG Developers Conference** or entering the following URL into your mobile browser: https://crowd.cc/s/1rKy0
Enter event code: **DevCon2018**

Alternatively, access here: https://crowd.cc/pcisig2018

Note: Create an account within the app so Admin knows who to contact if selected as the prize winner.

Each session feedback is provided is equivalent to 1 raffle entry (up to 11 sessions).
General survey feedback = 1 raffle entry.

Copyright © 2018 PCI-SIG® - All Rights Reserved